



## Algorithmic Refinements to an Enhanced Poisson-Boltzmann Approach Used in BioMolecular Simulation

P. Kar, M. Seel, U. H. E. Hansmann, S. Höfinger

published in

*From Computational Biophysics to Systems Biology (CBSB07),  
Proceedings of the NIC Workshop 2007,*  
Ulrich H. E. Hansmann, Jan Meinke, Sandipan Mohanty,  
Olav Zimmermann (Editors),  
John von Neumann Institute for Computing, Jülich,  
NIC Series, Vol. **36**, ISBN 978-3-9810843-2-0, pp. 173-176, 2007.

© 2007 by John von Neumann Institute for Computing

Permission to make digital or hard copies of portions of this work for personal or classroom use is granted provided that the copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise requires prior specific permission by the publisher mentioned above.

<http://www.fz-juelich.de/nic-series/volume36>

# Algorithmic Refinements to an Enhanced Poisson-Boltzmann Approach Used in BioMolecular Simulation

Parimal Kar<sup>1</sup>, Max Seel<sup>1</sup>, Ulrich H. E. Hansmann<sup>1,2</sup>, and Siegfried Höfner<sup>1</sup>

<sup>1</sup> Michigan Technological University,  
Department of Physics,  
1400 Townsend Drive, 49931 Houghton, MI, USA  
*E-mail: [pkar, seel, hansmann, shoefing]@mtu.edu*

<sup>2</sup> John von Neumann Institute for Computing, FZ Jülich  
52425 Jülich, Germany  
*E-mail: u.hansmann@fz-juelich.de*

In a series of recent publications we have introduced a very general description of solvation for biomolecular structure. It was shown that the enhanced continuum electrostatics approach can be decomposed into a series of individual terms, each of them representing its own portion of distinct physical interaction. Care has been taken to operate the model at conditions that guaranteed a maximum level of numerical accuracy. However, a number of internal parameters still need to be optimized further in order to speed up the procedure. Among these factors are i) the exact value of the exit criterion used to terminate the calculation, ii) the array dimension regulating the allowed number of consecutive DIIS steps, iii) the switch criterion used to move from the pre-DIIS stage to the DIIS stage, iv) the dependence on system size of the number of necessary iterations to achieve convergence, v) the dependence on renormalization factors applied to the net sum of polarization charges, vi) the influence of very small-sized boundary elements, or the introduced change when merging these very small-sized elements to larger ones from the neighborhood, and vii) the surface resolution necessary to calculate the dispersion term. Therefore in this present study we want to address these points and examine their consequences on run-time performance. A series of ten proteins of increasing size will be used as a testbed.

## 1 Introduction

Biological molecules typically reside in aqueous environments. Reliable consideration of the effect of water on structure and dynamics of biomolecules is among the key factors governing accurate descriptions of biological matter<sup>1</sup>. Here we focus on an implicit solvation model. Among other methods, e.g. SASA, GB, FDPB, the Poisson - Boltzmann (PB) approach<sup>2</sup> within the Boundary Element Method (BEM)<sup>3</sup> is frequently chosen due to its intermediate position regarding computational cost versus achievable accuracy. In our recent series of publications<sup>4,5</sup> we have outlined a generalization of the Polarizable Continuum Model (PCM)<sup>6</sup> applied to biomolecular structure. Each of the considered terms represents a separate portion of distinct physical interaction,

$$\Delta G^{sol} = \Delta G^{pol} + \Delta G^{disp,rep} + \Delta G^{cav} \quad (1)$$

which are polarization, dispersion and cavitation. The latter plays an important role in hydrophobicity related phenomena<sup>7</sup>. Care has been taken to operate the model at conditions

that guaranteed a maximum level of numerical accuracy. However, a number of internal parameters could still profit from further optimization.

## 2 Methods

### 2.1 Aims

In this present study, we address the following factors and examine their consequences on run-time performance with regard to a series of test proteins of increasing size that we have studied earlier<sup>5</sup>, (i) the exact value of the exit criterion used to terminate the calculation of the polarization term,  $\Delta G^{pol}$ , (ii) the array dimension regulating the allowed number of consecutive DIIS<sup>8</sup> steps, (iii) the switch criterion used to move from the pre-DIIS stage to the DIIS stage, (iv) the dependence on system size of the number of necessary iterations to achieve convergence, (v) the dependence on renormalization factors applied to the net sum of polarization charges, (vi) the influence of very small-sized boundary elements (BEs), or the introduced change when merging these very small-sized elements to larger ones from the neighborhood, (vii) the necessary degree of surface resolution for accurate calculation of the dispersion term,  $\Delta G^{disp}$ .

### 2.2 Procedure

We select 10 proteins of different size ( number of residues reaching from 41 to 430 ). Initially we run the PB/BEM program POLCH<sup>9</sup> at default conditions. The run time for all the 10 cases is recorded and forms a reference set. At first, we adjust the parameter MAXNIT which defines the maximum number of successive DIIS steps, hence determines the size of the DIIS matrix, and compute the run time deviation from the reference set for all the 10 proteins. Once parameter MAXNIT is optimized, we rerun the entire test set and extract net solvation free energies,  $\Delta G^{sol}$ , which serve as a new reference. ACCURA is the second parameter to be optimized. It defines the threshold criterion used for termination of the iterative process when computing the polarization term,  $\Delta G^{pol}$ . For optimizing ACCURA we require the deviation from the reference set not to exceed  $\pm 0.05$  kcal/mol for any of the proteins. Once ACCURA is optimized we redo the whole set of test proteins at optimized conditions for either parameter, ACCURA as well as MAXNIT. We extract the number of iterations needed for completion and use these as a new reference. In our next step we optimize the parameter DSNTRC. This parameter sets the switch criterion used to move from a pre-DIIS stage to the DIIS stage. It represents the mean square deviation of two successive sets of polarization charges. We keep changing DSNTRC and optimize for a minimum number of necessary iterations. The next point is concerned with renormalization of the polarization charges according to Gauss' Law. We study the effect this has on net solvation free energies. The solvation free energies obtained after renormalization form another reference set for our next investigation. Here, we study the influence of very small-sized BEs. We will merge these very small-sized elements to larger ones from the neighborhood. We change the parameter REQSZ (the required minimum size of a BE) and compute the deviation of solvation free energies from the reference set. We again do not allow the energy to change more than by  $\pm 0.05$  kcal/mol in all test runs. Finally, we use all previously optimized parameters for a final test focusing on the

Protein PDB Code	No. of Res.	Molecular Charge (a.u.)	No. of Iter.	$\Delta G^{sol}$ Without Normalization (kcal/mol)	$\Delta G^{sol}$ Including Normalization (kcal/mol)	$\Delta G^{sol}$ Deviation (unsigned) (kcal/mol)
1P9GA	41	+3	9	-319.73	-321.54	1.81
2B97	70	+2	8	-40.22	-39.26	0.96
1LNI	96	-5	10	-534.64	-536.39	1.75
1NKI	134	+5	10	-456.20	-454.94	1.26
1EB6	177	-11	10	-1224.17	-122.75	1.42
1G66	207	-2	9	-118.26	-119.01	0.75
1PIX	250	+3	11	-636.41	-636.41	0.00
1RTQ	297	-16	11	-1998.72	-2011.23	12.51
1YQS	345	+2	11	-217.22	-217.22	0.04
1GPI	430	-12	12	-1259.54	-1271.59	12.05

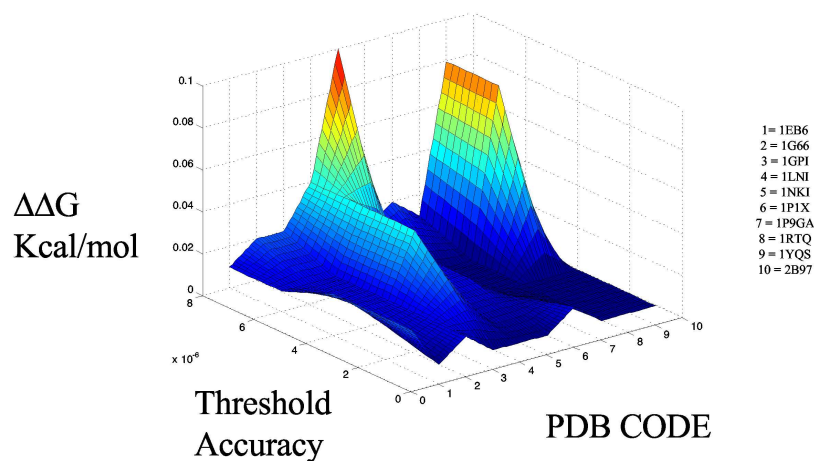


Figure 1. Numerical sensitivity of the employed enhanced Poisson-Boltzmann approach to the threshold criterion used for termination of the iterative sequence to calculate the polarization term,  $\Delta G^{pol}$ . A series of 10 proteins is tested and the threshold criterion is varied between  $1.0 \times 10^{-6}$  and  $8.0 \times 10^{-6}$ . When requiring the results to be numerically accurate at least up to the first digit behind the decimal point, ie allowing fluctuations  $< \pm 0.05$  (light blue patches on the surface) then the optimal value for this termination criterion is identified as  $4.0 \times 10^{-6}$ .

dispersion term. We change the resolution of the boundary used for calculation of  $\Delta G^{disp}$  which need not be maintained at such rigorous levels as identified for the polarization term<sup>4</sup>.

### 3 Results and Conclusions

Sensitivity to total system size, total charge and renormalization attempts is represented in Table 3. Variation of the termination criterion is graphically represented in Figure 1. In summary we find that the following parameters lead to a reasonable degree of numerical accuracy. (1) Best performance is achieved when the DIIS matrix is dimensioned  $7 \times 7$ , (2) Using a threshold criterion of  $4.0 \times 10^{-6}$  for termination of the iterative sequence occurring in  $\Delta G^{pol}$  computation leads to stable numerical results. (3) The best switch criterion to move from the pre-DIIS stage to the DIIS stage is given when the root mean square deviation between two successive sets of polarization charges falls below 0.05 a.u. (4) The number of iterations necessary to achieve convergence does not depend on system size. (5) A renormalization process will affect the net solvation free energies,  $\Delta G^{sol}$ , on the order of  $\pm 1-2\%$  of their total values. Systems with large net charges are more sensitive to renormalization. (6) If we merge small sized BEs to larger ones then no significant changes will occur when this procedure is limited to elements smaller than 8 % of the mean size ( $0.31 \text{ \AA}^2$ ). A reduction in number of BEs will lower the computational cost and foster numerical stability. (7) For calculation of the dispersion term,  $\Delta G^{disp}$ , we can reduce the discretization of the boundary into BEs of average size  $0.45 \text{ \AA}^2$  without loss of accuracy.

### Acknowledgments

This work was supported in part by the National Institutes of Health Grant GM62838.

### References

1. B. Honig, A. Nicholls, *Classical electrostatics in biology and chemistry*, Science **268**, 1144–1149, 1995.
2. J. Warwicker, H. C. Watson, *Calculation of the electric potential in the active site cleft due to alpha-helix dipoles*, J. Mol. Biol. **157**, 671–679, 1982.
3. R. J. Zauhar, R. S. Morgan, *A new method for computing the macromolecular electric potential*, J. Mol. Biol. **186**, 815–820, 1985.
4. P. Kar, Y. Wei, U. H. E. Hansmann, S. Höfinger, *Systematic Study of the Boundary Composition in Poisson Boltzmann Calculations*, J. Comput. Chem., in press, 2007.
5. P. Kar, M. Seel, U. H. E. Hansmann, S. Höfinger, *Dispersion Terms and Analysis of Size- and Charge-Dependence in an Enhanced Poisson-Boltzmann Approach*, submitted, 2007.
6. J. Tomasi, B. Mennucci, R. Cammi, *Quantum Mechanical Continuum Solvation Models*, Chem. Rev. **105**, 2999–3094, 2005.
7. M. Makowski, A. Liwo, H. A. Scheraga, *Simple physicsbased analytical formulas for the potentials of mean force for the interaction of amino acid side chains in water. I. Approximate expression for the free energy of hydrophobic association based on a Gaussian overlap model* J. Phys. Chem. B **111**, pp 2910–2916, 2007.
8. P. Pulay, *Convergence acceleration of iterative sequences. The case of scf iteration*, Chem. Phys. Lett. **73**, 393–398, 1980.
9. S. Höfinger, *Solving the Poisson-Boltzmann Equation with the Specialized Computer Chip MD-GRAPE-2*, J. Comp. Chem. **26**, 1148–1154, 2005.